ABSTRACT

A dictionary creation component [316] converts the normal orthographic written representation of a word into a sequence of symbols that relate to the pronunciation of the word. The symbols may be used to train conventional models for speech recognition. The symbols are not phonemes and do not need to be defined by a speech expert. The symbols are created automatically by the dictionary creation component based on the written representation of the word.